

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/123024/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Weiser, Rebecca ORCID: <https://orcid.org/0000-0003-3983-3272>, Green, Angharad E., Bull, Matthew J., Cunningham-Oakes, Edward, Jolley, Keith A., Maiden, Martin C.J., Hall, Amanda J., Winstanley, Craig, Weightman, Andrew J. ORCID: <https://orcid.org/0000-0002-6671-2209>, Donoghue, Denise, Amezcuita, Alejandro, Connor, Thomas R. ORCID: <https://orcid.org/0000-0003-2394-6504> and Mahenthiralingam, Eshwar ORCID: <https://orcid.org/0000-0001-9014-3790> 2019. Not all *Pseudomonas aeruginosa* are equal: strains from industrial sources possess uniquely large multireplicon genomes. *Microbial Genomics* , 000276. 10.1099/mgen.0.000276 filefile

Publishers page: <https://dx.doi.org/10.1099/mgen.0.000276>
<<https://dx.doi.org/10.1099/mgen.0.000276>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Not all *Pseudomonas aeruginosa* are equal: strains from industrial sources possess uniquely large multireplicon genomes

Rebecca Weiser^{1,†}, Angharad E. Green^{1,2,†}, Matthew J. Bull¹, Edward Cunningham-Oakes¹, Keith A. Jolley³, Martin C. J. Maiden³, Amanda J. Hall², Craig Winstanley², Andrew J. Weightman¹, Denise Donoghue⁴, Alejandro Amezcua⁵, Thomas R. Connor¹ and Eshwar Mahenthiralingam^{1,*}

Abstract

Pseudomonas aeruginosa is a highly versatile, antibiotic-resistant Gram-negative bacterium known for causing opportunistic infections and contamination of industrial products. Despite extensive genomic analysis of clinical *P. aeruginosa* strains, no genomes exist for preservative-tolerant industrial strains. A unique collection of 69 industrial isolates was assembled and compared to clinical and environmental strains; 16 genetically distinct industrial strains were subjected to array tube genotyping, multilocus sequence typing and whole-genome sequencing. The industrial strains possessed high preservative tolerance and were dispersed widely across *P. aeruginosa* as a species, but recurrence of strains from the same lineage within specific industrial products and locations was identified. The industrial *P. aeruginosa* genomes (mean=7.0 Mb) were significantly larger than those of previously sequenced environmental (mean=6.5 Mb; *n*=19) and clinical (mean=6.6 Mb; *n*=66) strains. Complete sequencing of the *P. aeruginosa* industrial strain RW109, which encoded the largest genome (7.75 Mb), revealed a multireplicon structure including a megaplasmid (555 265 bp) and large plasmid (151 612 bp). The RW109 megaplasmid represented an emerging plasmid family conserved in seven industrial and two clinical *P. aeruginosa* strains, and associated with extremely stress-resilient phenotypes, including antimicrobial resistance and solvent tolerance. Here, by defining the detailed phylogenomics of *P. aeruginosa* industrial strains, we show that they uniquely possess multireplicon, megaplasmid-bearing genomes, and significantly greater genomic content worthy of further study.

DATA SUMMARY

All newly determined *Pseudomonas aeruginosa* genome sequences have been deposited at the National Center for Biotechnology Information (NCBI) (BioProject PRJEB8749) with the sequence for strain RW109 (the largest industrial

strain in the dataset) available as GCA_900243355.1 (draft) and GCF_900243355.1 (complete). Additional draft *P. aeruginosa* genomes have been deposited as GCA_001374635.1, GCA_001373635.1, GCA_001373875.1, GCA_001374955.1, GCA_001374115.1, GCA_001374355.1, GCA_001374455.1,

Received 12 February 2019; Accepted 24 May 2019; Published 06 June 2019

Author affiliations: ¹Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Cardiff, Wales, UK; ²University of Liverpool, Institute of Infection and Global Health, Liverpool, UK; ³Department of Zoology, The Tinbergen Building, University of Oxford, Oxford, UK; ⁴Unilever Research and Development, Port Sunlight, Wirral, UK; ⁵Unilever Research and Development, Safety and Environmental Assurance Centre, Colworth House, Sharnbrook, Bedford, UK.

***Correspondence:** Eshwar Mahenthiralingam, mahenthiralingame@cardiff.ac.uk

Keywords: *Pseudomonas aeruginosa*; industry microbiology; contamination; phylogenomics; megaplasms.

Abbreviations: AMR, antimicrobial resistance; AT, Clondiag Array-Tube Typing; BA, benzoic acid; BIGSdb, Bacterial Isolate Genome Sequence database; BIT, benzisothiazolinone; CARD, Comprehensive Antibiotic Resistance Database; CF, cystic fibrosis isolate or strain; CGView, Circular Genome Viewer; CHX, chlorhexidine; CITMIT, chloromethylisothiazolinone; CLIMB, Cloud Infrastructure for Microbial Bioinformatics; CLIN, clinical isolate or strain (not CF); COG, cluster of orthologous group; ENV, environmental isolate or strain; GI, genomic island; GTR, generalised time-reversible; IND, industrial isolate or strain; IPARP, International *Pseudomonas aeruginosa* Reference Panel; ISO, International Organisation for Standardisation; iTOL, Interactive Tree of Life; MAFFT, Multiple Alignment using Fast Fourier Transform; MCR, module completeness ratio; MIC, minimum inhibitory concentration; MIT, methylisothiazolinone; MLST, multilocus sequence typing; NCBI, National Center for Biotechnology Information; PHASTER, PHAge Search Tool Enhanced Release; PHE, phenoxyethanol; QAST, Quality Assessment Tool for Genome Assemblies; RAPD, random amplified polymorphic DNA genotyping; rMLST, ribosomal multilocus sequence typing; SMRT, single molecule real-time; ST, sequence type; TSA, tryptic soya agar; TSB, tryptic soya broth; UPGMA, unweighted pair group method with arithmetic mean; WGS, whole genome sequencing.

†These authors contributed equally to this work

All supporting data, code and protocols have been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article.

000276 © 2019 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from www.microbiologyresearch.org by

IP: 131.251.254.243

On: Fri, 07 Jun 2019 10:45:53

GCA_001374655.1, GCA_001373655.1, GCA_001373895.1, GCA_001375215.1, GCA_001374135.1, GCA_001374375.1, GCA_001373595.1, GCA_001374675.1, GCA_001374995.1, GCA_001375235.1, GCA_001374155.1, GCA_001374395.1, GCA_001374975.1, GCA_001374435.1, GCA_001373675.1, GCA_001373915.1 and GCA_000568855.1 (full strain information is available in Table S1, available in the online version of this article)

INTRODUCTION

Pseudomonas aeruginosa is a widespread human opportunistic pathogen and has been highlighted as a key global threat in relation to antimicrobial resistance, as one of the ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *P. aeruginosa* and *Enterobacter cloacae*) pathogens [1]. *P. aeruginosa* is also responsible for chronic lung infections in people with cystic fibrosis (CF) [2], and the pathogenesis, epidemiology, treatment and antibiotic resistance of clinical strains have been major foci of research on the bacterium. However, *P. aeruginosa* is ubiquitous in the environment and can be isolated from diverse habitats, including soil, water and plants [3]. One significant but largely overlooked aspect of *P. aeruginosa*'s biology is its ability to survive in xenobiotic environments, such as industry, where it can contaminate raw materials, pharmaceutical and cosmetic products [4], and even fuels [5].

Within the home and personal care industry, microbial contamination is a major cause of product recalls and may represent a threat to consumer health [6]. *P. aeruginosa* has previously been linked with infections involving contaminated hand lotion [7], mouthwash [8], and make-up products [9]. Industrial manufacturers work to a number of guidelines and International Organization for Standardization (ISO) recommendations to prevent contaminated products from reaching the market place. Preservatives such as parabens, isothiazolinones, formaldehyde releasers, alcohols and organic acids are widely used to maintain product stability and prevent microbial proliferation [10]. When preservation barriers fail, global recall databases show that *P. aeruginosa* is a major cause of contamination, with, for example, >33 % of cosmetic product incidents involving the bacterium [11].

Despite the global reach and billion dollar value of the home and personal care industry, the study of *P. aeruginosa* strains that contaminate industrial products has been limited. What is clear is that these *P. aeruginosa* strains have the ability to overcome the harsh antimicrobial rich environments found in industry and survive to cause product recalls [4]. Understanding the biology of these micro-organisms is therefore important to protect consumers, prevent costly product contamination incidents, improve preservation strategies and assess if they are a reservoir for antimicrobial resistance. Industrial microbiology also represents a relatively unstudied area in terms of the application of genomics, despite microbial contamination impacting on multiple global manufacturing processes. Herein, we have assembled a unique collection of

Impact Statement

Multiple industrial products are prone to microbial contamination because they are made in a non-sterile way and rely on chemical preservatives to prevent spoilage. However, occasional microbial contamination can result in costly recalls and may pose health risks to consumers. We studied a unique collection of *Pseudomonas aeruginosa* bacteria recovered from industrial products. *P. aeruginosa* can cause infection in vulnerable people and is recognized as an antimicrobial resistance threat, but little is known about industrial strains. We DNA-sequenced the industrial strains and compared their genomes to those of clinical and environmental isolates. Industrial *P. aeruginosa* strains had high preservative tolerance and we found certain strain types reoccurring in the same product type or manufacturing location. Although the industrial strains were genetically diverse in relation to the *P. aeruginosa* population, we uniquely identified that they often carried the same megaplasmid. Megaplasms are large segments of DNA that can move between bacteria and carry new functions, such as antimicrobial resistance. The megaplasmid, plus other novel genes, gave the industrial *P. aeruginosa* strains very large genomes. What this extra capacity of 500+ genes brings to industrial *P. aeruginosa* in relation to its ability to overcome preservation represents a new challenge in microbiology.

P. aeruginosa industrial isolates, contextualized their genotypic and phenotypic characteristics against clinical and environmental strains, and revealed unique insights into their antimicrobial resistance, genome content and organization.

METHODS

Bacterial collection and phenotypic analysis

P. aeruginosa was recovered from 24 contaminated industrial products by serial dilution and growth on selective and non-selective agars as described previously [12]. The 69 industrial isolates were combined with a wider collection of 21 strains from the International *Pseudomonas aeruginosa* Reference Panel [13] (IPARP) and additional reference strains (Table S1). *P. aeruginosa* isolates were stored deep frozen, revived and routinely grown on tryptic soya agar (TSA) or in tryptic soya broth (TSB) (Oxoid Ltd, Basingstoke, UK) as described previously [14]. Multiple databases were analysed for strain information as follows: genome comparison, the *Pseudomonas* genome database [15] and GenBank; random amplified polymorphic DNA (RAPD) genotyping [16]; AT-typing and *EBURST* analysis [17]; multilocus sequencing typing (MLST), rMLST (ribosomal MLST) and whole-genome MLST [Bacterial Isolate Genome Sequence Database (BIGSdb)

[18]. Details of all 103 strains examined and which analyses were performed on each of them are given in Table S1.

Growth curve analysis and swimming, swarming and twitching motility assays were performed as described previously [14]. Preservative susceptibility testing of *P. aeruginosa* was carried out using a microbroth dilution assay essentially as described previously [19] and the minimal inhibitory concentration (MIC) was determined for six widely used industrial agents: chloromethylisothiazolinone (CITMIT; Kathon CG, Dow Europe GmbH, Switzerland); methylisothiazolinone (MIT; neolone M10, Dow Europe GmbH, Switzerland); benzisothiazolinone (BIT; Koralone B-120, Dow Europe GmbH, Switzerland); phenoxyethanol (PHE; Clariant Produkte GmbH, Germany); chlorhexidine (CHX; chlorhexidine digluconate; Sigma-Aldrich Co. Ltd, UK); and benzoic acid (BA; Sigma-Aldrich Co. Ltd, UK) (see Methods S1, Supplementary Material for full details).

A biofilm formation assay was adapted from a previously published protocol [20]. Suspensions of 10^5 colony-forming units (c.f.u.) ml^{-1} of each strain were prepared in TSB and 100 μl was inoculated into clear 96-well plates, 6 wells per strain (avoiding edge wells). After incubation at 37 °C for 32 h, growth was discarded and the wells were washed three times with sterile H_2O before staining with 200 μl 0.1 % (w/v) crystal violet solution. The wells were then washed with sterile H_2O until the washes were clear and left to air-dry overnight. The crystal violet in each well was solubilized with 200 μl 70 % ethanol and the absorbance measured at 570 nm. Three biological replicates were performed for each strain.

RAPD AND AT GENOTYPING OF *P. AERUGINOSA*

RAPD-PCR typing was performed as described previously using primer 272 [5]. RAPD-PCR products were separated using an Agilent 2100 Bioanalyzer and DNA 7500 chips according to the manufacturer's instructions. Cluster analysis of the RAPD-PCR profiles was performed using the Gel Compar package within Bionumerics 6.6 software (Applied Maths, Gent, Belgium). A Pearson correlation similarity coefficient with an unweighted pair group method with arithmetic mean (UPGMA) dendrogram was used to determine similarity between profiles; strains sharing ≥ 80 % profile similarity were considered to be the same RAPD type [5].

The Clondiag Array Tube (AT) *P. aeruginosa* genotyping kit (Alere Technologies GmbH, Jena, Germany) was used to genotype selected industrial/reference strains (Table S2) as described previously [21]. The relatedness of the *P. aeruginosa* strains was determined by comparison of the hexadecimal codes obtained for each strain to a 1464-strong database of clinical and environmental strains (those with complete profiles in the eBURST V3 database [17]) using the default algorithm (<http://eburst.mlst.net>) [22].

Draft genome sequencing of *P. aeruginosa*

DNA was extracted from 19 *P. aeruginosa* strains (16 industrial strains and 3 reference strains used in industrial testing; Table S3) as described previously [23] (see Method S2, Supplementary Material) and subjected to genome sequencing. Five additional *P. aeruginosa* strains (four clinical, one environmental and one of unknown source) were also subjected to sequencing (Table S3). Genomic library preparation and sequencing were performed at the Oxford Genomics Centre of the Wellcome Trust Centre for Human Genetics. DNA sequencing was achieved using 100 bp paired-end Illumina HiSeq2000 technology. *De novo* assembly of genomes from the paired-end sequence data was performed using SPAdes Genome Assembler v3.5.0. Assembled genomic data in contigs were deposited in the PubMLST database, supported by BIGSdb software [18]. MLST sequencing types (STs) were determined using the PubMLST database. The genomes were deposited in BIGSdb, assigned accession numbers (Table S1; BioProject PRJEB8749) and scanned for the seven *P. aeruginosa* MLST loci [24] to identify allele types and STs. The details of other strains in the database sharing the same ST were recorded to determine the origins of the STs.

COMPLETE SEQUENCING OF *P. AERUGINOSA* RW109

Single molecule real-time (SMRT) sequencing (Pacific Biosciences, PacBio, CA, USA) of *P. aeruginosa* RW109 (ENA accession GCF_900243355.1) was performed by the Centre for Genomic Research, University of Liverpool, essentially as described previously [23]. All bioinformatic analysis of the data was carried out using a virtual machine hosted by the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) consortium [25] and full details of the sequencing and assembly processes are given in the Supplementary Material (Method S3). The assembled genome was analysed as follows: annotation was carried out using Prokka [26]; cluster of orthologous group (COG) was determined as described previously [27]; Kyoto Encyclopedia of Genes and Genomes (KEGG) function was assigned [28] and examined for enrichment in preservative-tolerant strains; phages were located using the PHAge Search Tool Enhanced Release (PHASTER) (v1.0) [29]; and genomic islands were predicted using IslandViewer (v4.0) [30] (see Supplementary Material, Method S4, S5 and S6).

Comparative genome analysis of RW109 was performed against the complete genome sequence FASTA files of the clinical *P. aeruginosa* reference strains UCBPP-PA14 (PA14) (accession no. CP000438) and PAO1 (accession no. AE004091). Each genome was annotated using Prokka and run through the Quality Assessment Tool for Genome Assemblies (QUAST), (v4.5.5) [31], prior to comparison against the RW109 strain. The Circular Genome Viewer (CGView) application and CGView source code [32] were used via the command line to convert Prokka-generated GenBank files into CGView XML files for the generation of an RW109 genome circular map. The coding sequences were coloured according to the assigned COG

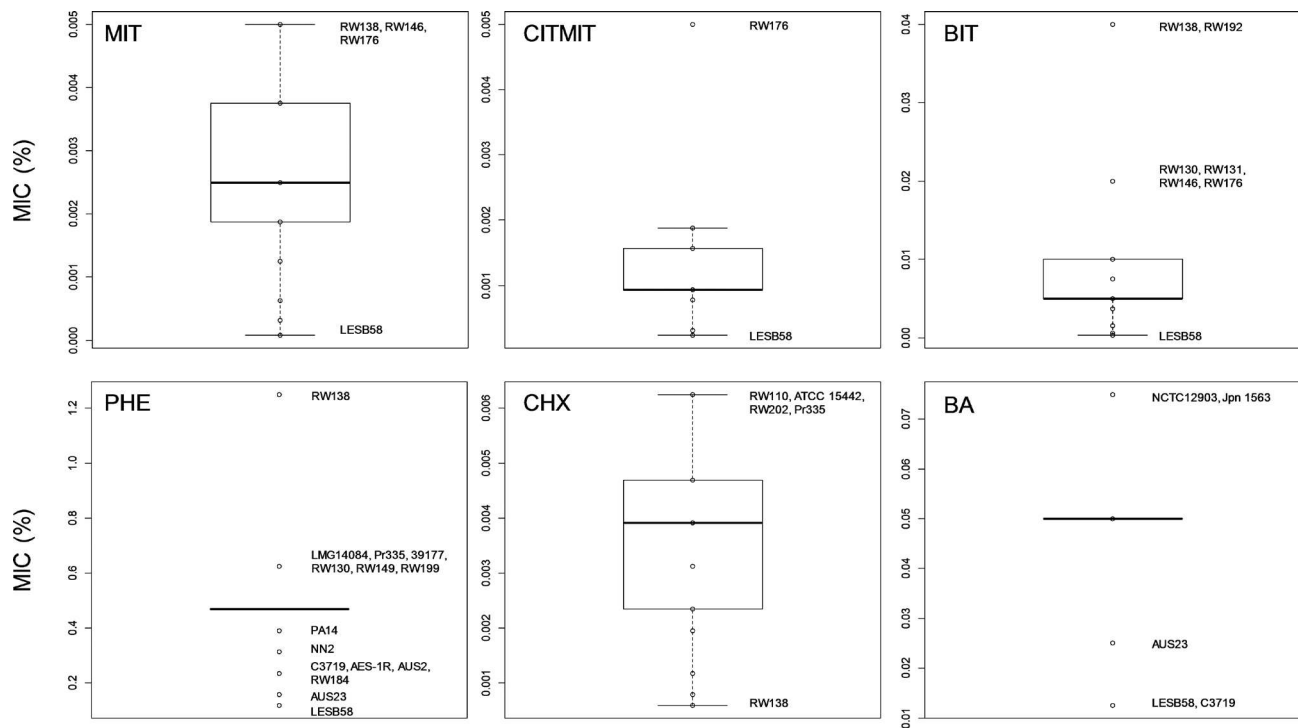


Fig. 1. *P. aeruginosa* demonstrates high tolerance of preservatives commonly used in industry. The minimum inhibitory concentrations (MICs) of preservatives were determined for the sub-panel of 40 *P. aeruginosa* strains. MIC data were visualized using boxplots and the median, upper quartile, lower quartile, maximum and minimum values are displayed. Data points are shown as open circles and outliers (minimum and maximum) are labelled with the strain name(s) possessing that MIC. Panels are labelled as follows: MIT, methylisothiazolinone; CITMIT, chloromethylisothiazolinone and methylisothiazolinone blend in a 3 : 1 ratio; BIT, benzisothiazolinone; PHE, phenoxyethanol; CHX, chlorhexidine; and BA, benzoic acid.

categories, and the RW109 GIs and prophage sequences were also illustrated on the map.

P. AERUGINOSA PAN-GENOME ANALYSIS

The Roary pipeline (v3.7.0) [33] was used for pan-genome comparisons of the RW109 strain with genome sequences of *P. aeruginosa* strains isolated from clinical, environmental and industrial sources. In total, a reference set of 103 genomes were analysed (Table S1). All the FASTA sequences of the strains were reannotated using Prokka [26] and the resulting GFF files were analysed with Roary [33]. The $-e$ parameter was applied to create a multi-FASTA codon-aware alignment of all the core genes using PRANK [34]. The core-gene multi-FASTA alignment file (MAFFT) was used to build a phylogenetic maximum-likelihood tree based on 1000 bootstrap resampling replicates with FastTree (v2.1.8) [35]. The generalized time-reversible (GTR) model was applied with default settings. The tree was visualized with FigTree (v1.4.3, <http://tree.bio.ed.ac.uk/software/figtree/>). From the Roary pan-genome analysis, the gene presence and absence output was used to determine the genes specific to the *P. aeruginosa* RW109 strain that were not identified in the 102 *P. aeruginosa* genome sequences.

Prediction of antimicrobial resistance genes

The FASTA sequences of the 103 *P. aeruginosa* genomes were analysed using the ABRicate tool (v0.5-dev, <https://github.com/tseemann/abrigate.git>) and antimicrobial resistance genes were predicted by the Comprehensive Antibiotic Resistance Database (CARD) [36]. An $\geq 80\%$ cut-off was used for both coverage and identity ABRicate scores. The ABRicate genome summary reports were combined into a single Excel spreadsheet to show the presence and absence of antimicrobial resistance genes. Visualization of the comparative antimicrobial resistance gene content was performed using the Interactive Tree of Life (iTOL) (v3.5.4) [37] by combining the spreadsheet and core-gene phylogenetic tree and producing a heat map.

Analysis of the RW109 megaplasmid

The CGView Comparison Tool [38] was used to map the *P. aeruginosa* industrial strains, along with the clinical strains PA121617 and PA96, to a circular map of the RW109 plasmid. Prokka-generated GenBank files were used as an input and were converted into CGView XML files for generation of the genome circular comparison map. For evolutionary analysis of the plasmid origin, the *parB* from the RW109 megaplasmid was used to extract homologous sequences by BLAST from GenBank.

Table 1. Comparative genome features of the industrial *P. aeruginosa* strain RW109 and the reference strains PA14 and PAO1

Genome features	Industrial strain RW109	Reference strains	
		PA14	PAO1
Genome size (bp)	7756224	6537648	6264404
Replicons	3	1	1
DNA G+C content (%)	65.11	66.29	66.56
Coding sequences	7303	5901	5671
No. hypothetical proteins	2199	2026	1858
rRNA genes	12	12	12
tRNA genes	76	69	73

A maximum-likelihood phylogenetic tree was generated based on MAFFT alignment of RW109 megaplasmid *parB*, homologous sequences identified in seven industrial and two clinical *P. aeruginosa* strains, and their closest neighbouring *parB* genes from characterized *Pseudomonas* plasmids [39] and other species. The accession numbers of the sequences analysed were as follows: the RW109 megaplasmid 1 (LT969519.1) and plasmids of the same ancestral family, pBM413 (CP016215.1), pSY153-MDR (KY883660.1), p12969-DIM (KU130294.1), pRBL16 (CP015879.1), pJB37 (KY494864.1), pTTS12 (NZ_CP009975.1) and pOZ176 (KC543497.1), and more distantly related plasmids, pTC-F14 (AF325537.2), AMMD p1 (CP000443.1), pAOVO02 (CP000541.1), Rms149 (AJ877225), pRA2 (U88088.2), pWW0 (AJ344068.1) and pUnnamed1 (NZ_CP029091.1; used as the phylogenetic root). The megaplasmid and large plasmid copy numbers were estimated by mapping the short read sequence data to the complete PacBio *P. aeruginosa* RW109 reference sequence using the EDGE bioinformatics platform [40]; the fold coverage of each replicon was compared to that of the main chromosome to derive the copy number estimate.

RESULTS

Industrial *P. aeruginosa* tolerate preservatives but are diverse in relation to other phenotypes

A collection of 69 *P. aeruginosa* isolates (Table S1) was assembled from industrial products (household cleaners, laundry liquids, personal care cosmetics, metal working fluids and timber care products) that were identified as not meeting hygiene standards during quality control checks of products manufactured between 2001 to 2013; reference strains used in preservative efficacy testing by industry were also included. Low-resolution genotyping (RAPD) was used to differentiate sequential isolates and identify single strains for further analysis (Fig. S1). Representative genetically distinct industrial strains (IND; $n=15$) were assembled into a testing panel of 40 *P. aeruginosa* strains from CF ($n=10$), other clinical infections (CLIN; $n=9$), environmental sources (ENV; $n=5$) and unknown ($n=1$) sources. Strains from IPARP [13] were also included (Table S1). Tolerance towards the following widely used preservatives was evaluated: isothiazolinones (methylisothiazolinone, MIT;

chloromethylisothiazolinone and methylisothiazolinone blended in a 3 : 1 ratio, CMIT; benzisothiazolinone, BIT), alcohols (phenoxyethanol, PHE), biguanides (chlorhexidine, CHX) and organic acids (benzoic acid, BA) (Fig. 1, Tables S4 and S5).

The industrial strains of *P. aeruginosa* demonstrated high levels of tolerance towards individual preservatives and individual industrial isolates possessed the maximum preservative MIC values for all classes (Fig. 1, Tables S4 and S5). Higher isothiazolinone MIC values (≥ 0.00375 % MIT, ≥ 0.001875 % CITMIT and ≥ 0.02 % BIT) were predominantly linked with industrial strains (Table S4). *P. aeruginosa* strains from industrial sources were found to have significantly higher ($P<0.05$) median preservative MICs than strains from CF for the isothiazolinones (Tables S4 and S5; MIT, 0.0025 compared to 0.001875; CITMIT, 0.0015626 compared to 0.0007815; BIT, 0.01 compared to 0.005; and PHE, 0.4688 compared to 0.2734). Other phenotypes, such as growth rate (Table S6) and motility (Table S7), were variable across the industrial strains. Only one strain, the industrial isolate RW200, demonstrated a true swarming phenotype, indicated by the formation of finger-like projections radiating from the inoculation point (Fig. S2a). In addition, three strains of industrial origin sharing the same genotype (RW130, RW131 and RW146; see below) isolated from a household cleaning product at the same location over several years (2004–2010), displayed the same unusual swimming motility (Fig. S2b).

The ability of the industrial *P. aeruginosa* strains to form biofilms was also investigated [20]. For the 16 industrial product isolates there were no associations between biofilm-forming capability (Fig. S3) and decreased susceptibility to the preservatives MIT, CITMIT, BIT, PHE, CHX or BA (Table 1). Biofilm formation across the 16 industrial, 4 CF, 4 clinical, and 2 environmental strains assessed was variable, but showed concordance with published data in terms of high and low biofilm-forming control strains that were included [14]. No association between strain source or planktonic antimicrobial tolerance and biofilm formation was established (Fig. S3), which was in agreement with what has been observed for *P. aeruginosa* [14] and *Burkholderia cepacia* complex bacteria [41].

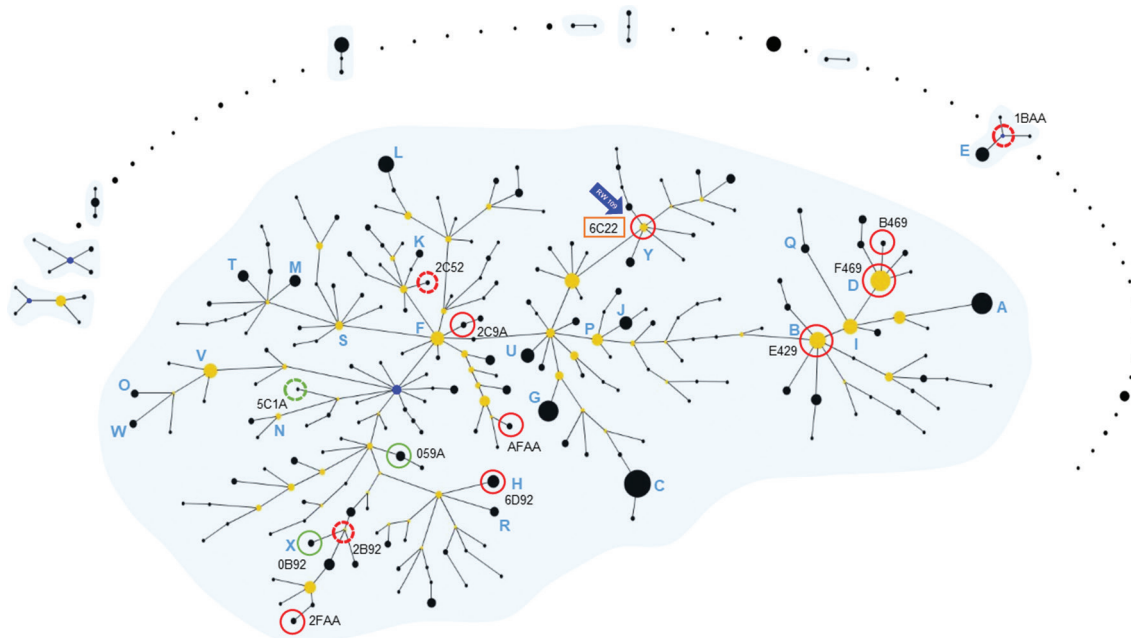


Fig. 2. Industrial *P. aeruginosa* strains are widely distributed across the AT genotype inferred population. The clonal complex structures of 1464 strains are shown, including strains from CF, other human infections and environmental sources, as well as 1446 strains analysed by Wiehlmann *et al.* [17]. The industrial and efficacy testing *P. aeruginosa* strains (Table S2) were analysed and placed within the inferred population as follows. Clonal complexes were calculated from the 16-marker AT genotype of the core genome by the EBURST algorithm; the major *P. aeruginosa* clonal complex and 8 additional groups have been highlighted with a blue background. AT genotypes are indicated by black, blue and yellow circles, with the diameter of the circle being proportional to the number of isolates with this genotype. Founding AT genotypes are shown in blue, subfounders are in yellow and prevalent clones are marked with light blue letters next to their corresponding AT genotype. The industry-associated *P. aeruginosa* strains are marked with closed (previously reported AT genotypes) or broken (novel AT genotypes) red circles, and their AT genotypes are indicated (Table S2). The position of the industrial strain RW109 is highlighted by the blue arrow and the orange box-enclosed AT genotype.

Industrial strains are widely spread across the non-clonal, freely recombining *P. aeruginosa* population structure

Multiple genotyping methods of increasing resolution were employed to examine the placement of industrial strains within the non-clonal recombinant *P. aeruginosa* population structure. A total of 19 representative industrial isolates (Fig S1; 16 industrial product isolates and 3 reference strains used in industry) were subjected to Clondiag AT genotyping [17] to enable comparison to the largest dataset of global strains currently available. AT genotypes were obtained for 15 industrial product isolates, revealing 11 different strain types (Table S2). To place these strains in a wider context, EBURST [22] was applied to the 1464 isolates [17] comprising the industrial genotypes. The 252 genotypes identified within this collection grouped into 1 large and 8 small clonal complexes, and 44 singletons (Fig. 2). The EBURST topology was highly similar to that observed by Cramer and colleagues [42], indicating a diverse spread of strains, with strains from different isolation sources being spread throughout the *P. aeruginosa* population (Fig. 2). In terms of specific AT genotypes, 3 of the 11 (27 %) of industrial genotypes were novel, with the remaining 8 strain types found being previously associated with varied

habitats, such as CF, other clinical settings and the natural environment, and previously well-documented *P. aeruginosa* clone types (major clone types Y, D, B and X).

Whole-genome sequencing (WGS) and MLST were used to obtain further information about *P. aeruginosa* population biology [18, 43]. MLST profiles [24] were generated for the industrial strains and represented 10 STs composed of known STs ($n=7$) and novel STs ($n=3$) (Table S3). Comparison to the global MLST database showed that the seven known STs associated with the industrial strains had been encountered in multiple environments, including CF, clinical infections and the environment (Table S3). ST-2730 was associated with three industrial strains (RW130, 131 and 146) isolated from the same household cleaner production facility from 2004 to 2010 (Table S3). This persistent *P. aeruginosa* industrial strain had a high tolerance towards benzoic acid (Fig. 1) and also possessed an atypical swimming motility (Fig. S2). Single industrial *P. aeruginosa* genotypes were also encountered in more than one industrial product type, with ST-111 occurring in two different household cleaning product types (isolates RW192 and RW138), and ST-1342 being encountered within a personal care cosmetic (isolate RW149) and laundry liquid (isolate RW176) product (Table S3).

High-resolution population biology mapping of the *P. aeruginosa* industrial strains was revealed by core-gene analysis of an expanded genome collection comprising 103 selected sequences (Table S1). A total of 4009 core genes were identified within this dataset, aligned and evaluated phylogenomically (Fig. 3). *P. aeruginosa* split into 2 major groups, with 13 industrial strains in group 1 and 3 in group 2 (Fig. 3). Three industrial strains possessing the same MLST sequence type (ST-111), RW138, RW109 and RW192 (Table S3), were located in a grouping within group 1, and were all differentiated by core gene analysis (Fig. 3). The industrial strains RW149, RW176 and RW200 were also located within this sub-clade, which only comprised industrial and clinical strains ($n=6$ for both; Fig. 3). Strains from different sources were widely dispersed across the core gene population analysis (Fig. 3), corroborating the AT genotyping (Fig. 2). The overall population structure of the two major groups and a PA7 outlier reflects the findings of studies with both smaller ($n=55$) [44] and larger ($n=389$ strains) [45] datasets.

Industrial *P. aeruginosa* strains have large multireplicon genomes

To establish a complete reference genome of an industrial *P. aeruginosa* strain, single-molecule real-time sequencing (Pacific Biosciences) was applied to strain RW109, which had the largest genome (Table 1, Fig. S4). This strain had originally been isolated from a product contamination incident (personal care cosmetic product, Europe, 2003; Table S1) and was closely related to five other industrial strains within the same sub-group of group 1 *P. aeruginosa* (Fig. 3). In comparison to the well-characterized clinical reference strains, *P. aeruginosa* PA01 and PA14 [46], industrial strain RW109 possessed: (i) a substantially larger genome of 7 756 224 bp; (ii) 1402 more predicted coding sequences than PA14 (the larger of the two well-studied reference strains, PA14 and PA01); and (iii) a multireplicon genome (Fig. 4) comprising a large chromosomal (7 049 347 bp; GC content 65%) replicon, a megaplasmid (555 265 bp; GC content 58 %; plasmid 1) and a large plasmid (151 612 bp; GC content 57 %; plasmid 2) (Table 1, Fig. S4). Other features, such as G+C, rRNA and tRNA content, were similar among the three strains (Table 1). Overall, the highly virulent *P. aeruginosa* PA14 strain genome was 15.7 % smaller, and the older PA01 reference genome was 19.2 % smaller, than that of the industrial strain RW109.

Comparative genome analysis of the *P. aeruginosa* industrial product strains revealed multiple unique traits. Firstly, genes homologous to those encoded by the RW109 megaplasmid were extensively conserved in other genetically distinct industrial strains (Fig. 4a). Comparative genomic analysis of the RW109 megaplasmid demonstrated that seven industrial strains (RW130, RW131, RW146, RW172, RW176, RW184 and RW204) encoded DNA with considerable synteny across >250 kbp of this replicon (Fig. 4a). Two clinical strains, PA121617 and PA96, also encoded DNA that was homologous to the RW109 megaplasmid (Fig. 4a). Collectively, these *P. aeruginosa* strains harbouring DNA homologous to the RW109 megaplasmid were spread widely across the

species population structure (seven distinct MLST sequence types; Fig. 3, Table S3). Secondly, evolutionary comparison of the plasmid-specific partitioning gene, *parB* [47], showed that the RW109 megaplasmid was from a single conserved family (Fig. 4b). The RW109 megaplasmid *parB* clustered with the *parB* genes from the same seven industrial strains and two clinical strains with the homologous gene content (Fig. 4a). The close phylogenetic relationship of the *P. aeruginosa* RW109 megaplasmid to characterized plasmids such as pJB37 and pOZ176 [48, 49] indicated that it belonged to the Inc-P2 incompatibility group. By mapping the short-read data to the PacBio complete reference sequence, the megaplasmid copy number was also estimated to be less than 2 (see Supplementary Material, Result S1), which is consistent with the previously reported low copy number associated with Inc-P2 megaplasmids [48, 49]. Overall, from the genomic (Fig. 4a) and phylogenetic (Fig. 4b) conservation, the data suggested that the ancestral backbone of the RW109 megaplasmid had been acquired by different host strains, but was a specific feature of *P. aeruginosa* industrial strains (Fig. 3).

The third trait shared by all industrial *P. aeruginosa* strains was a larger than average genome size for the species (Fig. 5). The industrial strain genomes were on average 5.7 % (>380 000 bp) larger than *P. aeruginosa* strains from other sources (Fig. 5), despite their genetic heterogeneity at the strain level (Fig. 3). Examination of the genome size distribution for all 103 *P. aeruginosa* revealed a mean size of 6 706 516 bp. Five *P. aeruginosa* industrial strains, RW109, RW146, RW130, RW131 and RW172, were outliers possessing the largest genomes in the dataset, with only one other strain, the clinically derived Carb01-63 encoding similar overall DNA content (Fig. 5a). The distribution of genome sizes in relation to strain source (Fig. 5b, c) demonstrated that overall *P. aeruginosa* recovered from industrial settings possessed significantly larger genomes, with a mean size of 7 040 410 bp, compared to clinical strains (mean, 6 660 367 bp; $P=0.0012$) and environmental strains (mean, 6 516 837 bp; $P=0.0002$; two-sample *t*-test assuming unequal variance). In the case of the industrial *P. aeruginosa* strains RW109, RW130, RW131, RW146, RW172, RW176, RW184 and RW204, the presence of the conserved megaplasmid was likely a contributing factor for this increased overall genome size (Fig. 4a, b).

Industrial *P. aeruginosa* strains have equivalent overall antimicrobial resistance gene content

Antimicrobial resistance (AMR) genes were predicted from 103 *P. aeruginosa* genomes using the CARD [50]. Visualization of the predicted AMR genes demonstrated that industrial strains encoded a similar range of antibiotic resistance genes to *P. aeruginosa* strains from other environments (Fig. 6). The mean number of CARD resistance genes was determined as 43 for CF and clinical strains, 41 for environmental strains and 41 for industrial strains. All 103 genome sequences had the polymyxin resistance genes *pmrB* and *arnA* and the aminoglycoside resistance gene *OXA-50*. Within the industrial genome sequences, RW109 and RW192 were the only two with the beta-lactamase PDC-1 gene; this gene was also found within

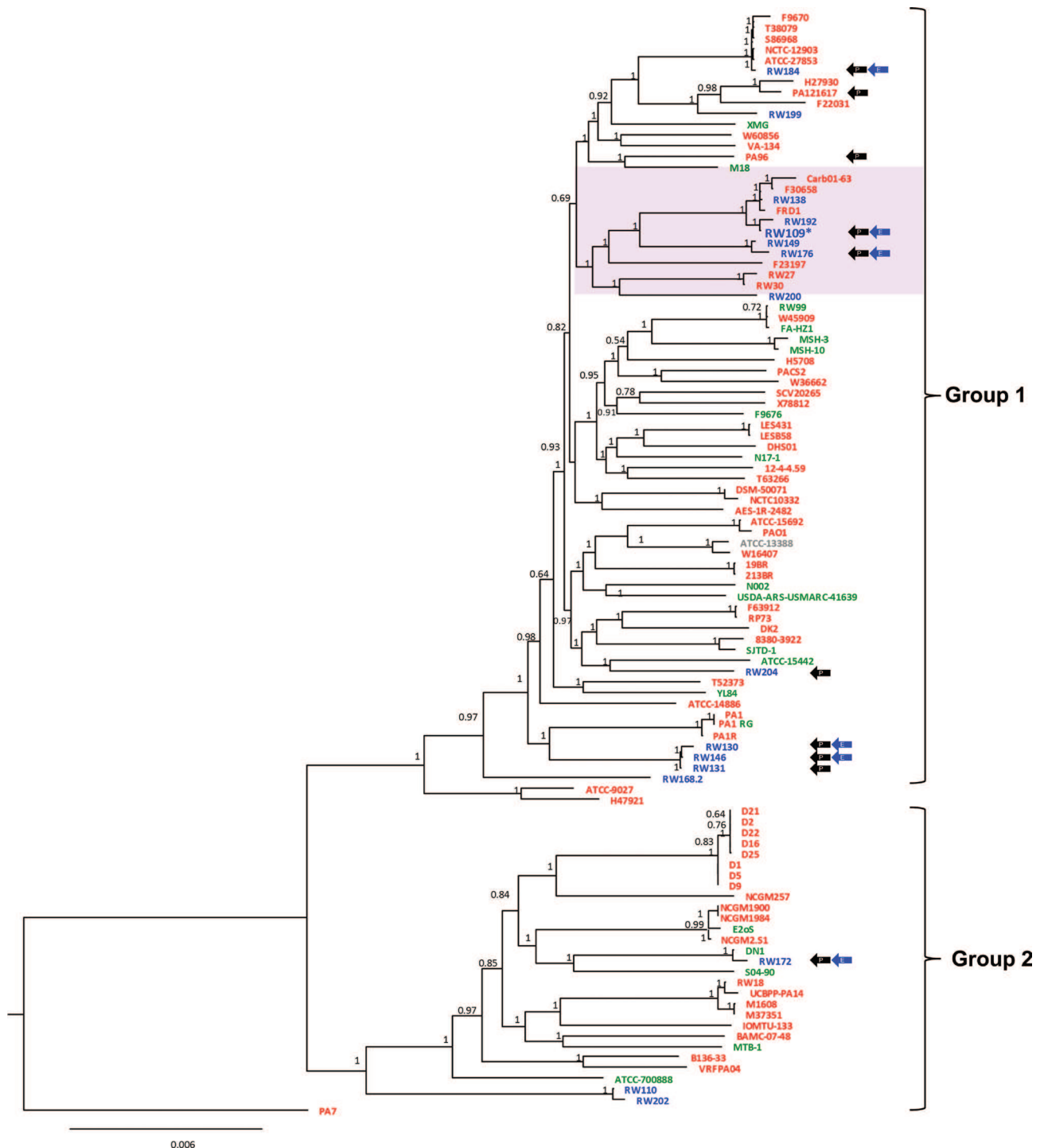


Fig. 3. Widespread placement of industrial isolates within the core gene-based population structure of *P. aeruginosa*. A maximum-likelihood phylogenetic tree was generated based on the 409 core genes identified within the 103 *P. aeruginosa* genomes. The strain names are colour coded to indicate their source: clinical (red, $n=66$), environmental (green, $n=19$) or industrial (blue, $n=16$); strain PA1RG is classified as both clinical and environmental and so is coloured red and green, and ATCC-13388 is coloured in grey as it was isolated from an unknown source. Bootstrap values (with 1 being 100 %) and genetic distance (the number of base substitutions per site) are indicated on the nodes and scale bar, respectively. The tree was rooted with the PA7 genome sequence as the most divergent *P. aeruginosa* strain. The two major *P. aeruginosa* groups are labelled (1 and 2), and the highlighted clade (purple square) containing the industrial strain, RW109, with a complete reference genome, is shown. Industrial strains uniquely encoding the BpeEF-OprC efflux pump (blue arrows; E) and the megaplasmid *parB* gene (black arrows; P) are indicated.

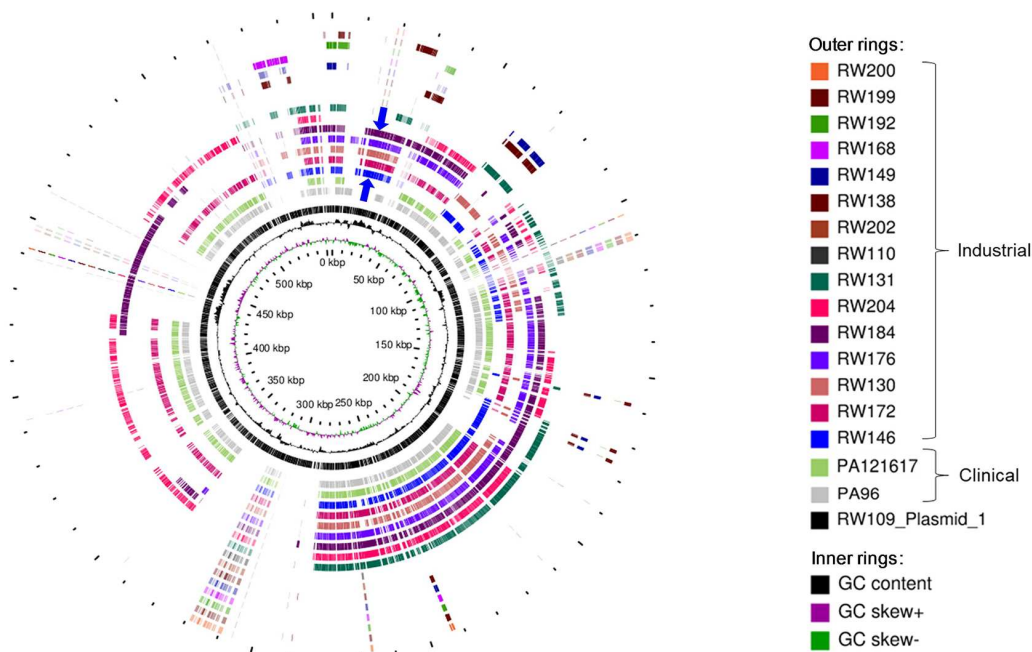
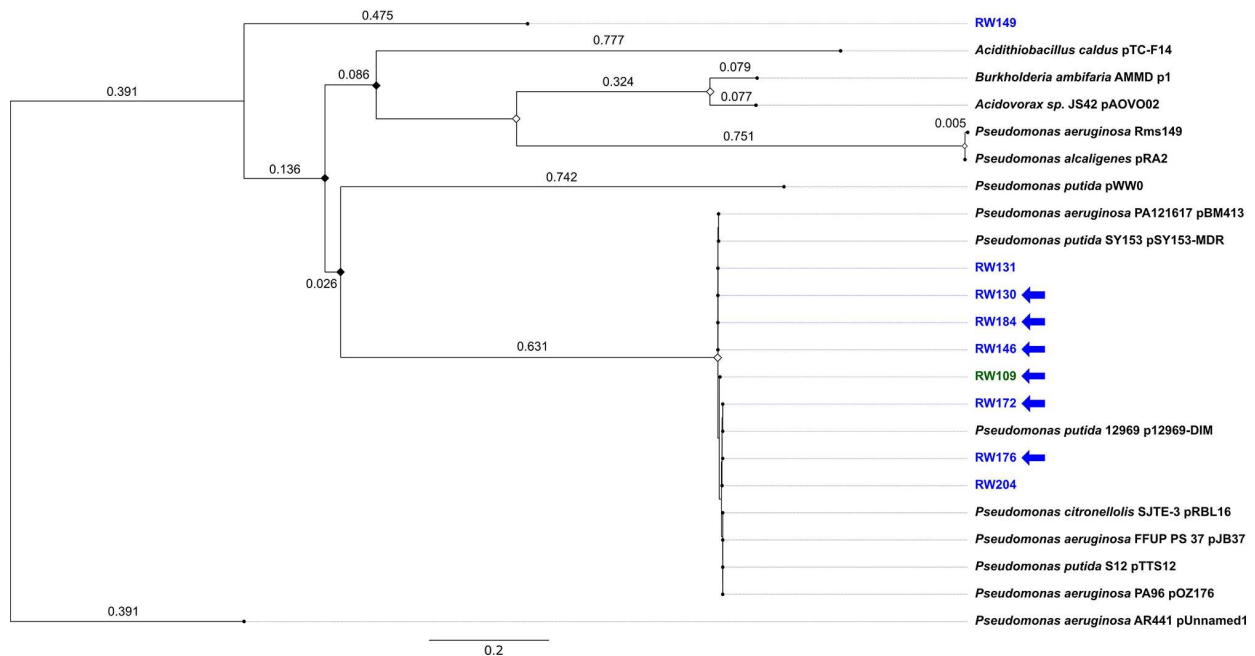
(a) Homology to other *P. aeruginosa* strain genomes(b) Conservation of *parB*

Fig. 4. Homology of the RW109 megaplasmid and its conservation among extremely stress-tolerant *Pseudomonas* strains. (a) The CGView Comparison Tool [38] was used to map the *P. aeruginosa* industrial strain sequences, along with the clinical strains PA121617 and PA96, to a circular map of the RW109 megaplasmid. Prokka-generated GenBank files were used as an input and were converted into CGView XML files for generation of the genome circular comparison map. The colour-coded key shows regions of each strain with homology to the megaplasmid, with their source indicated by the parentheses. The blue arrows indicate the approximate location of the conserved BpeEF-OprC efflux pump present in the industrial strains. (b) A maximum-likelihood phylogenetic tree of the RW109 megaplasmid *parB* gene (green text) aligned against homologous sequences from the *P. aeruginosa* industrial strains (blue text) and other related *Pseudomonas* plasmids. Industrial strains uniquely encoding components of the BpeEF-OprC efflux pump are indicated by the blue arrows. Bootstrap values are indicated at the node (diamonds; white nodes $\geq 75\%$ and black nodes $\leq 75\%$) and the genetic distance (the number of base substitutions per site) is indicated on the branches and by the scale bar. The tree was rooted with the *parB* gene from an unnamed plasmid in *P. aeruginosa* strain AR441.

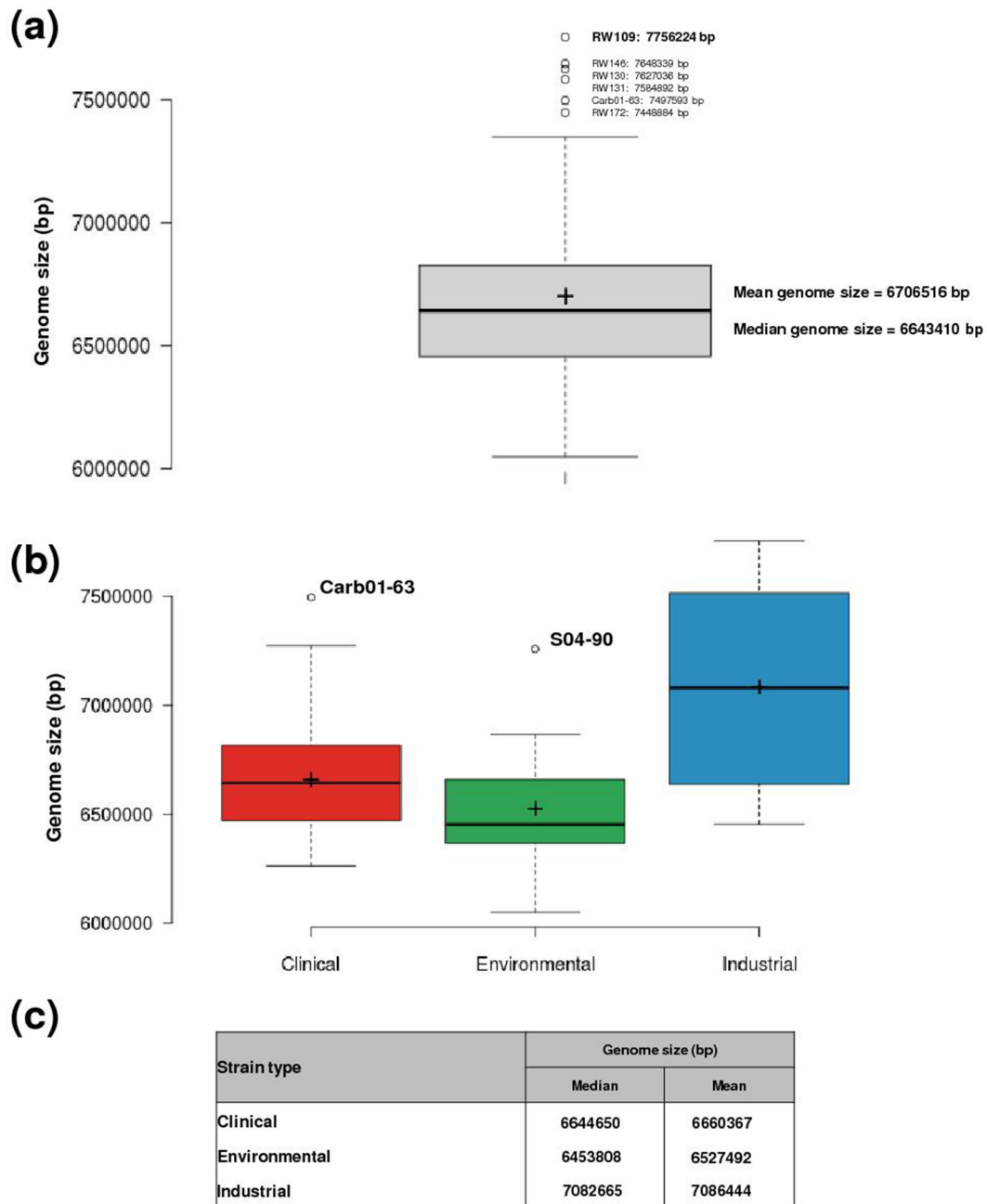


Fig. 5. *P. aeruginosa* strains from industrial product contamination possess considerably larger genomes than those from other sources. The distribution of the genome sizes for the 103 *P. aeruginosa* sequences were visualized using boxplots that display the median (black horizontal line), upper quartile, lower quartile, maximum and minimum values. The mean values are also represented with a cross. Outliers are shown as open circles outside of the boxplots, with the strain names and genome sizes indicated. A boxplot to represent the genome size distribution of the 103 sequences combined are shown in (a). The sequences were also divided into three boxplots (b) to indicate the genome sizes of the clinical strains (red, $n=66$), environmental strains (green, $n=19$) and industrial strains (blue, $n=16$). The values for the median and mean genome sizes (bp) for the clinical, environmental and industrial strains are shown in (c).

seven clinical strains (Fig. 4; ATCC-15692, H27930, H5708, PA121617, PAO1, VA-134 and W16407). CARD did not identify any antimicrobial resistance genes associated with the quinolone, carbapenem, trimethoprim, macrolide and tetracycline efflux resistance functional groups. The majority of the genes that encode the Mex-type efflux pump systems

were found in all *P. aeruginosa* industrial strains; exceptions were *oprA* (absent from all industrial strain genomes) and the *mexY* gene, which was specifically absent in RW109.

Pan-genome and presence/absence analysis was performed with the 103 *P. aeruginosa* genome dataset to determine

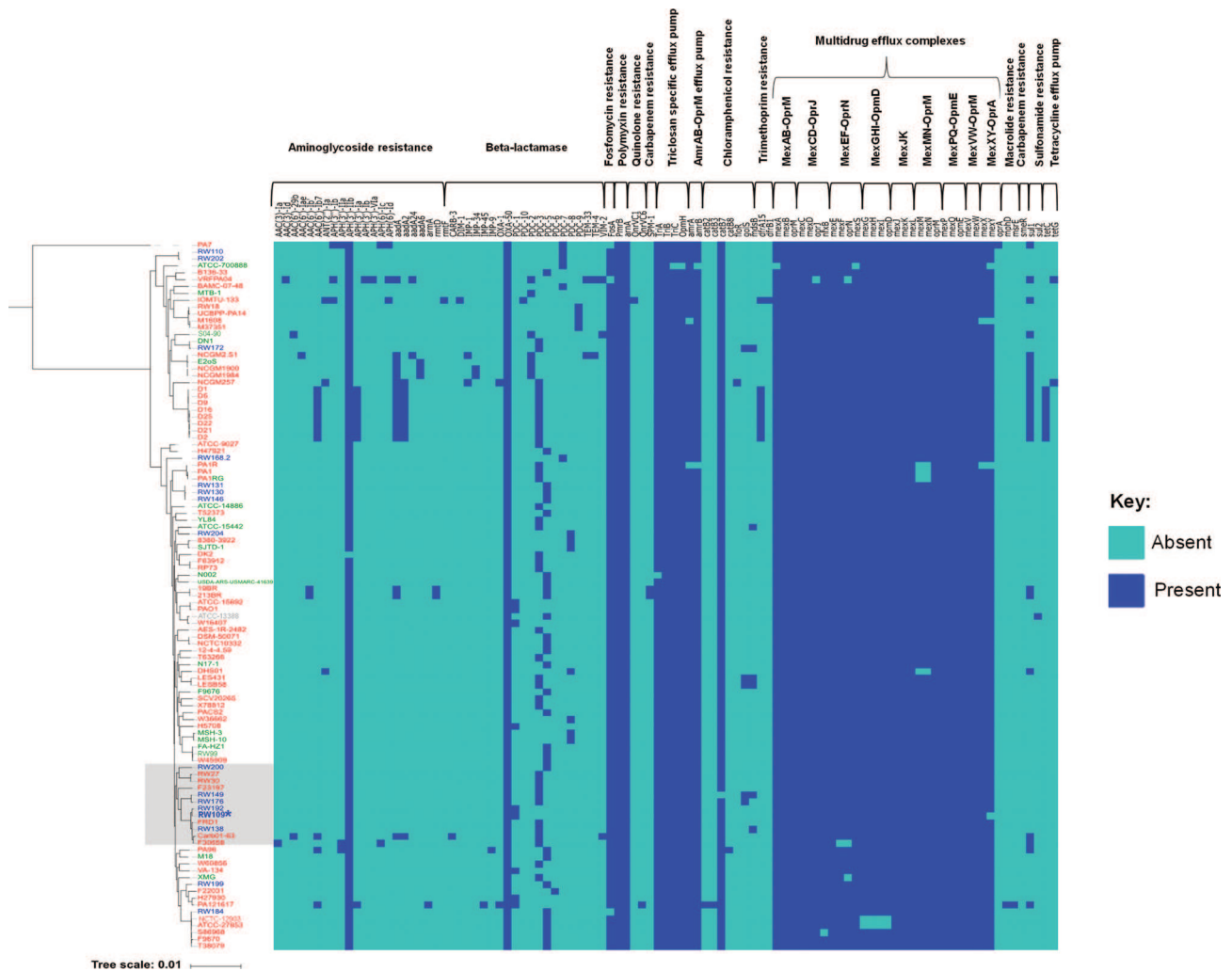


Fig. 6. Predicted antimicrobial resistance gene content of the 103 *P. aeruginosa* genomes. Using core-gene phylogenomic analysis to place each *P. aeruginosa* genome in evolutionary context, a presence/absence heat map was used to visualize the CARD-predicted antimicrobial resistance gene content. *P. aeruginosa* genome sequences of clinical (red), environmental (green) and industrial (blue) strains are indicated. The CARD antimicrobial genes were grouped via their predicted resistance functions (top of the heat map), with light blue showing the absence and dark blue indicating the presence of the antimicrobial resistance gene. The clade containing the industrial reference strain, RW109, for which a complete genome was determined, is highlighted (purple box).

genes unique to RW109. A total of 329 genes were found to be specific to the industrial strain, with 202 encoded on the main chromosome, 105 on megaplasmid 1 and 22 on plasmid 2. Over half (51%) of the RW109 unique genes were identified as poorly characterized by COG annotation assessment. No CARD-predicted antimicrobial resistance genes were found within the unique subset of RW109 genes. However, five genes associated with a BpeEF-OprC type efflux pump [51] identified by KEGG analysis [module M00698, module completeness ratio (MCR) of 100 % and Q-value of 0) were encoded by the RW109 megaplasmid (genes RW109_00010 through RW109_00014). The industrial strains RW130, RW146, RW172, RW176 and RW184, all of which likely possessed a megaplasmid from the same family of as that of RW109 (Fig. 4), also encoded the same KEGG-predicted efflux pump

genes (MCR of 100 % and a Q-value of 0). Genes homologous to the BpeEF-OprC type efflux pump were not identified in any of the remaining 97 *P. aeruginosa* genome sequences. The industrial strains possessing the BpeEF-OprC efflux pump genes were genetically distinct and spread across both major groups of the *P. aeruginosa* population structure (Fig. 3).

KEGG functional module pathway analysis was carried out on four isolates with the highest MICs and four with the lowest MICs for the preservatives BIT, MIT, CITMIT, PHE and CHX (see Fig. 1; Tables S4 and S5). Overall, there were minimal differences in the numbers of complete modules assigned to the categories between the isolates with high and low preservative MICs (Figs S5 and S6). Significant differences were associated with a limited number of functional module

pathways, including drug resistance, central carbohydrate metabolism, bacterial secretion system and two-component regulatory system categories (see Supplementary Results; Figs S5 and S6).

DISCUSSION

Researchers commonly portray *P. aeruginosa* as ubiquitous, linking this descriptor to its diverse phenotypic and genotypic traits, and multiple isolation sources. As an opportunistic pathogen, *P. aeruginosa* has been the subject of extensive investigation, but outside of clinical settings it has received much less attention despite its problematic nature when encountered. With the number of consumer products including preservatives increasing [52], and the production volume of biocidal compounds being several orders of magnitude higher than that of antibiotics [53], industrial manufacturing environments represent a key niche for stress-tolerant organisms such as *P. aeruginosa*. However, in contrast to infection, where *P. aeruginosa* isolates are frequently genotyped and archived to enable global epidemiological analysis, no systematic collections of *P. aeruginosa* isolates from industrial contamination have been assembled. By establishing a unique collection of *P. aeruginosa* isolates from industrial sources spanning a decade, different global locations and multiple product types, we have shown that they possess multireplicon genomes encoding 300 kb of extra DNA compared to strains from other sources.

Preserved industrial products represent an environment where antimicrobial compounds may exist within a concentration gradient ranging from lethal to sub-lethal concentrations [52], and such conditions place huge selective pressure on potential contaminating micro-organisms such as *P. aeruginosa*. Understanding the fitness advantage that *P. aeruginosa* gains from retaining genomic and megaplasmid content for survival within industrial environments is vital to develop improved preservation strategies and formulations. The International *Pseudomonas aeruginosa* Consortium is sequencing over 1000 genomes to study genome evolution, antibiotic resistance and virulence genes [45]. The initial findings from this project examining the core genomic content of 389 strains identified 3 major phylogenetic groups: group 1, which contained well-studied reference strains such as PAO1 [54] and LESB58 [55]; group 2, where the virulent reference strain PA14 [56] resided; and group 3, containing isolate PA7 [57] as an outlier. Our core gene analysis of a representative set of 103 *P. aeruginosa* genomes mirrored this analysis, and moreover has shed light on the widespread dispersal of industrial strains within both major phylogenetic groups (Fig. 3).

The consensus population structure of *P. aeruginosa* is non-clonal epidemic, composed of freely recombining isolates without habitat selection, but with evidence of globally distributed major clone types such as clone C and PA14 [3]. Confounding this model are potential niche specialists in clinical and environmental settings [3, 58–60]. The largest *P. aeruginosa* population structure analysis to date used AT genotyping to examine 2192 isolates from 1448 independent

habitats and included 8 isolates from oil-contaminated environments, but did not evaluate representative strains from contaminated industrial products [17]. We found that these industrial *P. aeruginosa* strains were diverse and a mixture of novel and previously encountered AT genotypes (Fig. 2). A recent small-scale study of animal infection as an under-reported niche of *P. aeruginosa* found the same diversity of genotypes within carbapenem-resistant strains [61]. The industrial AT genotypes were spread widely throughout the topology of the EBURST analysis (Fig. 2), corroborating the non-clonal epidemic hypothesis of *P. aeruginosa* population structure [42]. Furthermore, investigating the antimicrobial resistance gene content did not provide an obvious genetic link to the increased preservative tolerance of the industrial strains (Fig. 1), as strains of different origin shared similar numbers of core genome antimicrobial resistance determinants (Fig. 6). Sharing of resistance capabilities and a lack of clonal specialization in relation to antimicrobial phenotypes has previously been reported for *P. aeruginosa* strains from different environments [62].

Despite this population mixing and multiple ubiquitous traits being shared among *P. aeruginosa* strains, our genomic analysis of industrial strains demonstrated specialization in the direction of greater genomic capacity and multiple replicon genomes. It is interesting that *Burkholderia*, well known for their multireplicon genome structure, are also highly stress-tolerant and intrinsically antimicrobial-resistant [63]. A key feature of the reference industrial genome for *P. aeruginosa* strain RW109 defined herein was the presence of a 555 kbp megaplasmid (Fig. 4). Megaplasmids are well known to encode a wide range of accessory fitness traits that contribute to specialist ecological lifestyles [47]. The RW109 megaplasmid origin was conserved within seven other industrial strains and its gene content was homologous to megaplasmids within *Pseudomonas* species recovered from a wide variety of stressful environments (Fig. 4a, b). These *Pseudomonas* megaplasmids (all >370 kbp) have each been characterized individually for their fitness-contributing roles in clinical multidrug resistance for pSY153-MDR, pBM413, p12969-DIM, pOZ176 [64] and pJB37 [48]; mediating biodegradation processes such as oestrogen and aromatic hydrocarbon breakdown in wastewater treatment (pRBL16) [65]; and solvent tolerance and the production of aromatic hydrocarbons (pTTS12) [66]. The *Pseudomonas putida* megaplasmid, pTTS12 (583 kbp) is the largest characterized for the genus to date [66], and the RW109 megaplasmid (555 kbp) uniquely identified in this study is the largest associated with a strain of *P. aeruginosa*. Recent analysis of *P. aeruginosa* isolates recovered from patients at a hospital in Thailand also showed that the same closely related family of megaplasmids encode multidrug resistance elements, are highly recombinogenic and mediate extensive gene transfer, and have likely been carrying resistance genes dating back to the late 1970s [67].

Specific characterization of this unique megaplasmid family in relation to the wider fitness they bring to *Pseudomonas*, as well as their role in preservative tolerance within industrial isolates, is worthy of further work. Our initial analysis has

demonstrated no clear association of megaplasmid with *P. aeruginosa* phenotypes such as growth rate (Table S6), motility (Table S7) or biofilm formation (Fig. S3), but overall the industrial strains harbouring the plasmid showed elevated levels of preservative tolerance (Fig. 1). The KEGG functional module content was also broadly the same for the comparison of the limited numbers of well-characterized preservative-tolerant and -sensitive *P. aeruginosa* strains we were able to compare (Figs S5 and S6). It was interesting that the bacterial secretion category was enriched in isothiazolinone-tolerant *P. aeruginosa* strains (Fig. S5), and that the BpeEF-OprC type efflux pump was an underpinning pathway in this functional content difference (Supplementary Material, Result S2). Systematic comparison of large numbers of *P. aeruginosa* strains and functional analysis of the implicated gene pathways will be required to fully validate these initial correlations. While projects such as the 1000 genomes study [45] are delivering extensive sequence information, more data for poorly studied phenotypes such as preservative tolerance are required to perform high-resolution comparative analysis of *P. aeruginosa*.

CONCLUSIONS

P. aeruginosa is rightly described as an opportunistic pathogen, but the data presented here point towards the fact that it is actually a highly opportunistic organism, possessing genomic flexibility that enables it to take advantage of new niches wherever it can find them. The current study has uniquely characterized a collection of *P. aeruginosa* strains of industrial origin that are impacting on global product manufacture and has increased our understanding of this important bacterial species. *P. aeruginosa* is a common contaminant in xenobiotic settings and we identified problematic strains with high preservative tolerance that were capable of causing recurrent industrial incidents. These findings suggested that industrial strains possessed elevated fitness levels and present a case for strain typing and tracking within manufacturing environments as part of quality control. Further investigation into preservative tolerance will inform preservative system development and evaluation, but will require appropriate model strains. The prototypic strains PAO1 and PA14 [46] had significantly smaller genomes and lower preservative tolerance than the industrial strains. We focused our attention on *P. aeruginosa* RW109 and defined a complete reference genome for this industrial strain, discovering that it possessed one of the largest genomes and a megaplasmid associated with the species. Whilst there may not be specialized clones within industry, understanding the potential fitness advantage of the extra genomic content and megaplasmids shared by *P. aeruginosa* industrial strains is a key future challenge.

Funding information

This work was supported by Biotechnology and Biological Sciences Research Council (BBSRC) awards to R. W. (BB/F016557/1), A. E. G. (BB/L502078/1/1) and E. C. -O. (BB/M009122/1), with additional industrial CASE funding from Unilever Research and Development, UK. T. R. C. and M. B. acknowledge funding from the Medical Research Council's

Cloud Infrastructure for Microbial Bioinformatics (MR/L015080/1), the computational platform we used for genomic analyses.

Author contributions

Conceptualization, E. M., D. D. and A. A.; methodology, R. W. and A. E. G.; software, R. W., A. E. G., M. J. B., T. R. C. and K. J.; validation R. W., A. E. G., E. C. O. and E. M.; formal analysis, R. W., A. E. G., M. J. B. and E. C. O.; investigation, R. W., A. E. G., M. J. B., E. C. O. and A. H.; resources, D. D., A. A., K. J., T. R. C., M. C. J. M., C. W. and E. M.; data curation, R. W., A. E. G., M. J. B., T. R. C. and K. J.; writing – original draft preparation, E. M., R. W. and A. E. G.; writing – review and editing, E. M., R. W., A. E. G., E. C. O., M. J. B., K. J. and T. R. C.; visualization, R. W., A. E. G. and E. C. O.; supervision, E. M., A. J. W., T. R. C., D. D. and A. A.; project administration, E. M., D. D. and A. A.; funding, E. M., T. R. C., D. D. and A. A.

Conflicts of interest

D. D. and A. A. are employees of Unilever Research and Development, the company that part-funded the Biotechnology and Biological Sciences Research Council (BBSRC) Industrial CASE PhD awards to R. W., A. E. G. and E. C. -O. The remaining authors have no conflicts of interest to declare.

Data bibliography

1. Weiser RW, Bull MJ, Jolley KA, Maiden MCJ and Mahenthiralingam E. European Nucleotide Archive, PRJEB8749 (2015).

References

1. Pendleton JN, Gorman SP, Gilmore BF. Clinical relevance of the ESKAPE pathogens. *Expert Rev Anti Infect Ther* 2013;11:297–308.
2. Lipuma JJ. The changing microbial epidemiology in cystic fibrosis. *Clin Microbiol Rev* 2010;23:299–323.
3. Pirnay JP, Bilocq F, Pot B, Cornelis P, Zizi M *et al*. *Pseudomonas aeruginosa* population structure revisited. *PLoS One* 2009;4:e7740.
4. Jimenez L. Microbial diversity in pharmaceutical product recalls and environments. *PDA J Pharma Sci Technol* 2007;61:383–399.
5. White J, Gilbert J, Hill G, Hill E, Huse SM *et al*. Culture-independent analysis of bacterial fuel contamination provides insight into the level of concordance with the standard industry practice of aerobic cultivation. *Appl Environ Microbiol* 2011;77:4527–4538.
6. Lundov MD, Moesby L, Zachariae C, Johansen JD. Contamination versus preservation of cosmetics: a review on legislation, usage, infections, and contact allergy. *Contact Derm* 2009;60:70–78.
7. Stephenson JR, Heard SR, Richards MA, Tabaqchali S. Outbreak of septicaemia due to contaminated mouthwash. *Br Med J* 1984;289:1584–1584.
8. Becks VE, Lorenzoni NM. *Pseudomonas aeruginosa* outbreak in a neonatal intensive care unit: a possible link to contaminated hand lotion. *Am J Infect Control* 1995;23:396–398.
9. Reid FR, Wood TO. *Pseudomonas* corneal ulcer: the causative role of contaminated eye cosmetics. *Arch Ophthalmol* 1979;97:1640–1641.
10. Steinberg DC. Regulatory review: 2010 frequency of preservative use. *Cosmetics and Toiletries* 2010;125:46–51.
11. Neza E, Centini M. Microbiologically Contaminated and Over-Preserved Cosmetic Products According Rapex 2008–2014. *Cosmetics* 2016;3:3.
12. Weiser R, Donoghue D, Weightman A, Mahenthiralingam E. Evaluation of five selective media for the detection of *Pseudomonas aeruginosa* using a strain panel from clinical, environmental and industrial sources. *J Microbiol Methods* 2014;99:8–14.
13. De Soyza A, Hall AJ, Mahenthiralingam E, Drevinek P, Kaca W *et al*. Developing an international *Pseudomonas aeruginosa* reference panel. *Microbiologyopen* 2013;2:1010–1023.
14. Cullen L, Weiser R, Olszak T, Maldonado RF, Moreira AS *et al*. Phenotypic characterization of an international *Pseudomonas aeruginosa* reference panel: strains of cystic fibrosis (CF) origin show less in vivo virulence than non-CF strains. *Microbiology* 2015;161:1961–1977.
15. Winsor GL, Van Rossum T, Lo R, Khaira B, Whiteside MD *et al*. *Pseudomonas* Genome Database: facilitating user-friendly,

- comprehensive comparisons of microbial genomes. *Nucleic Acids Res* 2009;37:D483–D488.
16. Mahenthiralingam E, Campbell ME, Foster J, Lam JS, Speert DP. Random amplified polymorphic DNA typing of *Pseudomonas aeruginosa* isolates recovered from patients with cystic fibrosis. *J Clin Microbiol* 1996;34:1129–1135.
 17. Wiehlmann L, Cramer N, Tümmmler B. Habitat-associated skew of clone abundance in the *Pseudomonas aeruginosa* population. *Environ Microbiol Rep* 2015;7:955–960.
 18. Jolley KA, Maiden MCJ. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;11:595.
 19. Rushton L, Sass A, Baldwin A, Dowson CG, Donoghue D et al. Key role for efflux in the preservative susceptibility and adaptive resistance of *Burkholderia cepacia* complex bacteria. *Antimicrob. Agents Chemother.* 2013;57:2972–2980.
 20. O'Toole GA. Microtiter dish biofilm formation assay. *J Vis Exp* 2011;47.
 21. Wiehlmann L, Wagner G, Cramer N, Siebert B, Gudowius P et al. Population structure of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 2007;104:8101–8106.
 22. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 2004;186:1518–1530.
 23. Song L, Jenner M, Masschelein J, Jones C, Bull MJ et al. Discovery and biosynthesis of gladiolin: A *Burkholderia gladioli* antibiotic with promising activity against *Mycobacterium tuberculosis*. *J Am Chem Soc* 2017;139:7974–7981.
 24. Curran B, Jonas D, Grundmann H, Pitt T, Dowson CG. Development of a multilocus sequence typing scheme for the opportunistic pathogen *Pseudomonas aeruginosa*. *J Clin Microbiol* 2004;42:5644–5649.
 25. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J et al. CLIMB (the cloud Infrastructure for microbial bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2016;2:e000086.
 26. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
 27. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 2000;28:33–36.
 28. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007;35:W182–W185.
 29. Arndt D, Grant JR, Marcu A, Sajed T, Pon A et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–W21.
 30. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 2017;45:W30–W35.
 31. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
 32. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. *Bioinformatics* 2005;21:537–539.
 33. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S et al. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
 34. Löytynoja A. Phylogeny-aware alignment with PRANK. *Multiple sequence alignment methods* 2014:155–170.
 35. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;26:1641–1650.
 36. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA et al. The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 2013;57:3348–3357.
 37. Letunic I, Bork P. Interactive tree of life (iTOL) V3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242–W245.
 38. Stothard P, Grant JR, Van Domselaar G. Visualizing and comparing circular genomes using the CGView family of tools. *Brief Bioinform* 2017;7.
 39. Thomas CM, Haines AS, Koshleva IA, Boronin A. *Pseudomonas* plasmids. In: Rehm BHA (editor). *Pseudomonas Model Organism, Pathogen, Cell Factory*. Weinheim, Germany: Wiley-VCH GmbH & Co; 2008. pp. 293–330.
 40. Li PE, Lo CC, Anderson JJ, Davenport KW, Bishop-Lilly KA et al. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Res* 2017;45:67–80.
 41. Rose H, Baldwin A, Dowson CG, Mahenthiralingam E. Biocide susceptibility of the *Burkholderia cepacia* complex. *J Antimicrob Chemother* 2009;63:502–510.
 42. Cramer N, Wiehlmann L, Ciofu O, Tamm S, Høiby N et al. Molecular epidemiology of chronic *Pseudomonas aeruginosa* airway infections in cystic fibrosis. *PLoS One* 2012;7:e50731.
 43. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 2012;158:1005–1015.
 44. Stewart L, Ford A, Sangal V, Jeukens J, Boyle B et al. Draft genomes of 12 host-adapted and environmental isolates of *Pseudomonas aeruginosa* and their positions in the core genome phylogeny. *Pathog Dis* 2014;71:20–25.
 45. Freschi L, Jeukens J, Kukavica-Ibrulj I, Boyle B, Dupont M-J et al. Clinical utilization of genomics data produced by the international *Pseudomonas aeruginosa* consortium. *Front Microbiol* 2015;6:1036.
 46. Kukavica-Ibrulj I, Bragonzi A, Paroni M, Winstanley C, Sanschargin F et al. In vivo growth of *Pseudomonas aeruginosa* strains PA01 and PA14 and the hypervirulent strain LESB58 in a rat model of chronic lung infection. *J Bacteriol* 2008;190:2804–2813.
 47. Petersen J. Phylogeny and compatibility: plasmid classification in the genomics era. *Arch Microbiol* 2011;193:313–321.
 48. Botelho J, Grosso F, Quinteira S, Mabrouk A, Peixe L. The complete nucleotide sequence of an IncP-2 megaplasmid unveils a mosaic architecture comprising a putative novel blaVIM-2-harboring transposon in *Pseudomonas aeruginosa*. *J Antimicrob Chemother* 2017;72:2225–2229.
 49. Xiong J, Alexander DC, Ma JH, Déraspe M, Low DE et al. Complete sequence of pOZ176, a 500-kilobase IncP-2 plasmid encoding IMP-9-mediated carbapenem resistance, from outbreak isolate *Pseudomonas aeruginosa* 96. *Antimicrob Agents Chemother* 2013;57:3775–3782.
 50. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 2013;57:3348–3357.
 51. Podnecky NL, Wuthiekanun V, Peacock SJ, Schweizer HP. The BpeEF-OprC efflux pump is responsible for widespread trimethoprim resistance in clinical and environmental *Burkholderia pseudomallei* isolates. *Antimicrob Agents Chemother* 2013;57:4381–4386.
 52. Gilbert P, McBain AJ. Potential impact of increased use of biocides in consumer products on prevalence of antibiotic resistance. *Clin Microbiol Rev* 2003;16:189–208.
 53. European Commission. Scientific committee on emerging and newly identified health risks: effects of the active substances in biocidal products on antibiotic resistance 2008.
 54. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrenner P et al. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 2000;406:959–964.
 55. Winstanley C, Langille MGI, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C et al. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the liver-pool epidemic strain of *Pseudomonas aeruginosa*. *Genome Res* 2009;19:12–23.

56. Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL et al. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol* 2006;7:R90.
57. Roy PH, Tetu SG, Larouche A, Elbourne L, Tremblay S et al. Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. *PLoS One* 2010;5:e8842.
58. Stewart RMK, Wiehlmann L, Ashelford KE, Preston SJ, Frimmersdorf E et al. Genetic characterization indicates that a specific subpopulation of *Pseudomonas aeruginosa* is associated with keratitis infections. *J Clin Microbiol* 2011;49:993–1003.
59. van Mansfeld R, Jongerden I, Bootsma M, Buiting A, Bonten M et al. The population genetics of *Pseudomonas aeruginosa* isolates from different patient populations exhibits high-level host specificity. *PLoS One* 2010;5:e13482.
60. Selezska K, Kazmierczak M, Müsken M, Garbe J, Schobert M et al. *Pseudomonas aeruginosa* population structure revisited under environmental focus: impact of water quality and phage pressure. *Environ Microbiol* 2012;14:1952–1967.
61. Haenni M, Hocquet D, Ponsin C, Cholley P, Guyeux C et al. Population structure and antimicrobial susceptibility of *Pseudomonas aeruginosa* from animal infections in France. *BMC Vet Res* 2015;11:9.
62. Valot B, Guyeux C, Rolland JY, Mazouzi K, Bertrand X et al. What it takes to be a *Pseudomonas aeruginosa*? the core genome of the opportunistic pathogen updated. *PLoS One* 2015;10:e0126468.
63. Mahenthiralingam E, Urban TA, Goldberg JB. The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat Rev Microbiol* 2005;3:144–156.
64. Yuan M, Chen H, Zhu X, Feng J, Zhan Z et al. pSY153-MDR, a p12969-DIM-related mega plasmid carrying bla IMP-45 and armA, from clinical *Pseudomonas putida*. *Oncotarget* 2017;8:68439–68447.
65. Zheng D, Wang X, Wang P, Peng W, Ji N et al. Genome sequence of *Pseudomonas citronellolis* SJTE-3, an estrogen- and polycyclic aromatic hydrocarbon-degrading bacterium. *Genome Announc* 2016;4.
66. Kuepper J, Ruijsenaars HJ, Blank LM, de Winde JH, Wierckx N. Complete genome sequence of solvent-tolerant *Pseudomonas putida* S12 including megaplasmid pTTS12. *J Biotechnol* 2015;200:17–18.
67. Cazares R, Moore MP, Grimes M, Emond-Rheault JG, Wright LL et al. A megaplasmid family responsible for dissemination of multi-drug resistance in *Pseudomonas*. *BioRxiv* 2019.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.